Erweiterte Policy-Version (mit Beispielen)

Titel

SVS-Prinzip: Strukturierte Trennung von Zustand und Bewertung in Entscheidungsprozessen

---

1. Ausgangslage

Moderne Entscheidungsumgebungen sind geprägt durch:

- hohe Unsicherheit
- komplexe Datenlagen
- zunehmenden Einsatz automatisierter Systeme

Ein strukturelles Problem besteht darin, dass:

> **Unsicherheit häufig direkt handlungswirksam wird.**

Dies führt zu:

- vorschnellen Entscheidungen
- Eskalation bei unklarer Lage
- systemischer Instabilität

---

2. Grundprinzip

SVS fordert:

> **Die konsequente Trennung von Zustandsbeschreibung und Bewertung.**

🔷 **Zustand (State)**

- beschreibt die rekonstruierte Realität
- Grundlage für operative Entscheidungen

🔷 **Bewertung (Valuation)**

- beschreibt Unsicherheit, Risiko, Vertrauen
- darf nur indirekt wirken

---

3. Entscheidungslogik

> **Handlungen basieren ausschließlich auf dem Zustand.**

Bewertung beeinflusst nur:

- Modus (z. B. Normalbetrieb vs. Vorsicht)

- Validierung (z. B. zusätzliche Daten)
- Überwachung

---

4. Beispiele

🏥 **Medizin**

- **Problem:** Unsichere Diagnose → aggressive Behandlung
- **SVS:** Unsicherheit → mehr Diagnostik, nicht mehr Therapie

👉 Ergebnis: weniger Überbehandlung

🤖 **KI-Systeme**

- **Problem:** Confidence-Schwankungen → instabile Entscheidungen
- **SVS:** Entscheidung stabil, Unsicherheit → Prüfung

👉 Ergebnis: robustere Systeme

⚙️ **Infrastruktur / Industrie**

- **Problem:** Messrauschen → Regelungsinstabilität
- **SVS:** Rauschen → Filterung, nicht Steuerung

👉 Ergebnis: stabilere Prozesse

🕊️ **Krisenmanagement**

- **Problem:** unklare Lage → Eskalation
- **SVS:** unklare Lage → Beobachtung und Rückkopplung

👉 Ergebnis: deeskalierendes Verhalten

---

5. Wirkung

SVS führt strukturell zu:

- Reduktion unnötiger Eingriffe
- Stabilisierung von Systemen
- Erhöhung der Entscheidungsqualität
- Verbesserung der Transparenz

---

6. Grenzen

SVS:

- verhindert keine Konflikte
- ersetzt keine politische Bewertung
- garantiert keine optimale Entscheidung

Aber:

reduziert systematisch Fehlreaktionen unter Unsicherheit

---

7. Leitsatz

Unsicherheit ist ein Signal für Vorsicht – nicht für Handlung.

📜 2. Normative Version (Policy / Standard)

2.1 Grundanforderung

MUST
Systeme müssen Zustand und Bewertung strikt trennen.

2.2 Entscheidungslogik

MUST
Operative Entscheidungen dürfen ausschließlich auf Zustandsinformationen basieren.

2.3 Verbot direkter Kopplung

MUST NOT
Bewertung darf keinen direkten Einfluss auf Handlungen haben.

2.4 Zulässige Nutzung von Bewertung

SHOULD
Bewertung sollte verwendet werden für:

- Moduswechsel
- zusätzliche Validierung
- konservative Anpassung
- Monitoring

2.5 Invarianz

SHOULD
Bei identischem Zustand sollten Entscheidungen unabhängig von Bewertung konsistent bleiben.

2.6 Sicherheit

MUST
Sicherheitsgrenzen müssen unabhängig von Bewertung und Lernprozessen definiert sein.

2.7 Auditierbarkeit

MUST
Systeme müssen nachweisen können:

- Trennung von Zustand und Bewertung
- kein direkter Pfad Bewertung → Handlung

2.8 Eskalationskontrolle

SHOULD
Unsicherheit sollte nicht zu unmittelbarer Eskalation führen.

⚙️ 3. Implementierungsleitfaden (praxisnah)

3.1 Architektonische Trennung

Systeme sollten in zwei Pfade getrennt werden:

- **Zustandspfad (entscheidungsrelevant)**
- **Bewertungspfad (meta-relevant)**

3.2 Designprinzipien

- keine gemeinsame Variable für Zustand und Unsicherheit
- keine direkten Trigger von Bewertung auf Aktion
- klare Schnittstellen zwischen Ebenen

3.3 Typische Umsetzung

🔷 Schritt 1: Zustand definieren

- robuste, gefilterte Repräsentation
- möglichst unabhängig von Unsicherheitsmetriken

🔷 Schritt 2: Bewertung isolieren

- Unsicherheitsmaße separat berechnen
- keine direkte Einspeisung in Entscheidungslogik

🔷 Schritt 3: Meta-Ebene definieren

- Regeln für:
  - Vorsicht
  - Validierung
  - Überwachung

🔷 Schritt 4: Sicherheitsmechanismen

- feste Grenzen (Hard Constraints)
- unabhängige Überprüfung

3.4 Prüfmechanismen

System sollte getestet werden auf:

- Verhalten bei hoher Unsicherheit
- Reaktion auf Rauschen
- Stabilität unter Stress

3.5 Zentrale Testfrage

> **Ändert sich die Handlung, wenn nur die Unsicherheit steigt, aber der Zustand gleich bleibt?**

- Ja → Fehler
- Nein → SVS-konform

⬅ END Gesamtfazit

SVS ist ein:

- **einfaches, aber tiefes Strukturprinzip**
- **technisch implementierbar**
- **ethisch anschlussfähig**
- **deeskalierend wirkend**

🧩 Finaler Kernsatz

Systeme sollten unter Unsicherheit vorsichtiger werden – nicht reaktiver.